

# Die Cambridge Structural Database: Rückblick und Vorausschau

Colin R. Groom\* und Frank H. Allen\*

CCDC · CSD · Kristallstrukturdatenbanken · Strukturaufklärung · Wirkstoffentwicklung

**D**as Cambridge Crystallographic Data Centre (CCDC) wurde 1965 gegründet. Das Ziel war die Einrichtung einer Institution, die numerische, chemische und bibliographische Daten von organischen und metallorganischen Kristallstrukturen archiviert. In der Cambridge Structural Database (CSD) sind heute nahezu 700 000 Strukturen hinterlegt, und sie ist ein historisch lückenloses Archiv der niedermolekularen Kristallographie. Fast 40 000 Neueinträge werden jedes Jahr verzeichnet. Da die Röntgenkristallographie ihr einhundertstes Jubiläum feiert – und das CCDC bald sein fünfzigstes –, wollen wir in diesem Kurzaufsatz die Anfänge des CCDC als eine öffentlich geförderte Institution und seine weitere Entwicklung in eine gemeinnützige Einrichtung nachverfolgen. Wir beschreiben die Entwicklung der CSD und ihres Softwaresystems und betrachten ihre Auswirkungen auf die Forschung in der Strukturchemie, den Materialwissenschaften sowie der Biologie und Medizin. Außerdem diskutieren wir das Finanzierungsmodell des CCDC vor dem Hintergrund der Open-Access- und Open-Data-Debatte.

## 1. Ursprünge der kristallographischen Datenbanken

Die Kristallographie wurde in jüngster Zeit reichlich mit Gedenkveranstaltungen gewürdigt, insbesondere zum 100-jährigen Jubiläum der ersten Kristallstrukturbestimmung (Zinkblende) durch W. L. und W. H. Bragg 1912 und 1913<sup>[1]</sup> sowie durch die Ausrufung des Internationalen Jahres der Kristallographie (IYCr) 2014 durch die Vereinten Nationen. Das IYCr bietet uns Gelegenheit, auf die Erfolge der Kristallographie zurückzublicken und zu untersuchen, wie Kristallstrukturdaten wichtige wissenschaftliche Entwicklungen beeinflusst haben und dies weiterhin tun. Die kristallographischen Datenbanken sind ein historisches Archiv der vergangenen hundert Jahre und waren für Wissenschaftler vieler

Disziplinen eine wichtige Quelle von Kristallstrukturinformationen. Was nun ist es, das Kristallstrukturdaten so unschätzbar wertvoll macht?

Nach den Arbeiten der Braggs hat sich die Röntgenstrukturanalyse von Kristallen sehr bald als eine ganz besondere Analysetechnik etabliert. Im Jahr 1929 – 16 Jahre nach der Entdeckung – begann man, die vereinzelt publizierten Ergebnisse von Kristallstrukturstudien in einer eigens gegründeten Zeitschrift zusammenzufassen, den *Strukturberichten*.<sup>[2a]</sup> Die *Strukturberichte* gingen nahtlos in die *Structure Reports*<sup>[2b]</sup> über, die bis in die 1990er Jahre als offizielle Publikation der International Union of Crystallography (IUCr) erschienen sind. 1929 war daher ein besonderes Jahr für die Kristallographie – und es war außerdem das Jahr, als Linus Pauling seine fünf Regeln für die Bestimmung der Strukturen komplexer anorganischer ionischer Kristalle veröffentlichte.<sup>[3]</sup> Diese Regeln waren aus der Analyse von Daten abgeleitet, die in den 16 Jahren zuvor gesammelt worden waren, und sie waren mit fast an Sicherheit grenzender Wahrscheinlichkeit das erste Beispiel von Strukturkorrelation oder Struktursystematik.

Die Kristallstrukturanalyse bewegte sich nun auch von ihren rein anorganischen Ursprüngen weg und wandte sich zusehends Problemen der allgemeinen Strukturchemie zu. Für organische Verbindungen bestätigte oder etablierte die Röntgenkristallographie grundlegende Erkenntnisse über die Natur der chemischen Bindung. Die Planarität des Benzolrings wurde bewiesen, die Strukturen einer immer größer werdenden Zahl von natürlichen und synthetischen Molekü-

[\*] Dr. C. R. Groom  
Executive Director, Cambridge Crystallographic Data Centre  
12 Union Road, Cambridge CB2 1EZ (Großbritannien)  
E-Mail: groom@ccdc.cam.ac.uk

Dr. F. H. Allen  
Emeritus Research Fellow, Cambridge Crystallographic Data Centre  
12 Union Road, Cambridge CB2 1EZ (Großbritannien)  
E-Mail: allen@ccdc.cam.ac.uk

len wurden bestimmt, und, was besonders wichtig war, es wurden intermolekulare Wechselwirkungen ausfindig gemacht, vor allem von Wasserstoffbrücken. All dieses Wissen sollte sich bald als ein Fundament für das Gebiet der Molekularbiologie erweisen, dessen Entwicklung eine gewaltige Herausforderung zu einer Zeit war, in der die Bestimmung selbst der kleinsten organischen Strukturen außerordentlich zeitaufwendig und schwierig war. Der Rest ist Geschichte – eine Geschichte, in der die Kristallographie auch eine wichtige Rolle bei der Vergabe von 28 Nobelpreisen gespielt hat.<sup>[4]</sup>

Die Gründung der *Strukturberichte* 1929 war ein erster Vorstoß der modernen computerisierten Kristallstrukturdatenbanken, die fast vierzig Jahre später ihren Betrieb aufnehmen sollten: Zuerst die 1965 gegründete Cambridge Structural Database (CSD; Cambridge, Großbritannien)<sup>[5]</sup> für organische und metallorganische Strukturen, dann in den 1970er Jahren die Inorganic Crystal Structure Database (ICSD; Karlsruhe),<sup>[6a]</sup> die Metals and Alloys Crystal Structures Database (CRYSTMET; Ottawa, Kanada),<sup>[6b]</sup> die Protein Data Bank<sup>[6c]</sup> (PDB; Brookhaven National Laboratory, USA, heute Worldwide PDB)<sup>[6d]</sup> und die Nucleic Acid Database (NDB; Rutgers University, USA).<sup>[6e]</sup>

## 2. Das Cambridge Crystallographic Data Centre

Zwischen 1929 und den 1960er Jahren waren gedruckte Datensammlungen die Bezugsquelle für Kristallstrukturdaten. Zu den beiden führenden Zeitschriften, den *Strukturberichten* und den *Structure Reports*, gesellten sich eine Reihe von spezialisierteren Sammlungen, z. B. *Crystal Data*,<sup>[7]</sup> *Tables of Interatomic Distances and Configuration in Molecules and Ions*<sup>[8]</sup> und andere. Während dieses Zeitraums wuchs bei Wissenschaftlern aller Disziplinen die Sorge darüber, wie man mit dem raschen Anstieg der wissenschaftlichen Literatur – der Informationsflut – Schritt halten könne. Internationale Tagungen wurden abgehalten, die erste war die Royal Society Scientific Information Conference in London 1948, die einen 700-seitigen Tagungsreport nach sich zog.<sup>[9]</sup> Eine zentrale Figur in dieser Anfangsphase war der Physiker und Kristallograph J. D. Bernal (1901–1971).<sup>[10]</sup>

Im Royal Society Conference Report<sup>[9]</sup> schrieb Bernal, dass „die wachsende Fülle wissenschaftlicher Veröffentlichungen und das Durcheinander beim Abfassen der Daten

einen Bremsklotz für den Fortschritt der Wissenschaft darstellt.“ Das Problem der Informationsflut wurde von vielen Wissenschaftlern und letztlich auch den Forschungsministerien vieler Länder erkannt. Man begann natürlich bald zu realisieren, welche Möglichkeiten die in der Entwicklung begriffene Computertechnologie bot. Eine gemeinsame Arbeitsgruppe der Royal Society und des britischen Wissenschaftsministeriums wurde gegründet, die die Aufgabe hatte, Großbritanniens Beitrag zu den weltweiten Anstrengungen bei der Entwicklung einer Datenbank zu koordinieren. Im Jahr 1964 wurde Olga Kennard, die mit Bernal am Londoner Birkbeck College an einigen der gedruckten Kompendien gearbeitet hatte,<sup>[7,8]</sup> eingeladen, ein „Crystallographic Data Centre“ mit Fördermitteln des neuen UK Office for Scientific and Technical Information (OSTI) zu gründen.

## 3. Die Cambridge Structural Database

Das Cambridge Crystallographic Data Centre (CCDC) wurde 1965 innerhalb des Department of Organic Chemistry an der University of Cambridge gegründet. Olga Kennard war eingeladen worden, eine Röntgenkristallographiegruppe zu gründen. Das CCDC erhielt die Aufgabe, eine vollständige und rückwirkende Computerdatenbank der organischen und metallorganischen, durch Beugungsmethoden (Röntgen und Neutronen) bestimmten Strukturen aufzubauen. Die Datenbank sollte von jeder einzelnen Strukturanalyse die bibliographischen, chemischen und kristallographischen Angaben sowie, was am wichtigsten war, die dreidimensionalen Atomkoordinaten enthalten. Während also computerbasierte bibliographische Datenbanken noch in der Anfangsphase ihrer Entwicklung waren, wurde das CCDC mit der Erschaffung der weltweit ersten vollelektronischen numerischen Datensammlung beauftragt. Wichtig war dabei auch, dass die CSD in einem wissenschaftlichen Institut mit der unmittelbaren Beteiligung aktiver Forscher entstand, was bis heute so geblieben ist.

Alle Informationen mussten anfangs aus gedruckten Zeitschriften in die CSD übertragen werden, einschließlich einem Rückbestand von 4000 Strukturen, der parallel zu der Aufnahme aktueller Publikationen in die Datenbank aufzuarbeiten war. Dies erforderte Systeme für die Identifizierung von Veröffentlichungen, welche Kristallstrukturen enthielten,



Colin R. Groom ist seit 2008 Executive Director am CCDC. Er promovierte in Proteinkristallographie an der University of Leeds und war Postdoktorand in Leeds und an der Massey University, Neuseeland. 1994 trat er bei Pfizer ein, wo er mit der Errichtung einer Proteinkristallographiegruppe betraut war, später leitete er Abteilungen für molekulare Informatik in Großbritannien und den USA. 2002 wechselte er zu UCB (Celltech), wo er verschiedene Forschungsgruppen leitete. Er ist Fellow der Royal Society of Chemistry und Mitglied im Kuratorium mehrerer Zeitschriften.



Frank H. Allen ist Emeritus Research Fellow am CCDC, wo er bis zu seiner Pensionierung 2008 Executive Director war. Er studierte Chemie und Kristallographie am Imperial College in London und war Postdoktorand an der University of British Columbia in Vancouver. 1970 trat er seine Stelle am CCDC an und war dort mit der Entwicklung der Cambridge Structural Database (CSD), Softwareentwicklungen und Anwendungen von CSD-Daten in der Forschung betraut. Er ist Fellow der Royal Society of Chemistry und erhielt die RSC Silver Medal und den Herman Skolnik Award. Er war von 1993–2002 Chefredakteur der *Acta Crystallographica*, Section B.

für die Aufbereitung der bibliographischen und chemischen Daten und für die manuelle Erfassung numerischer Datenreihen. Es erforderte auch neuartige Software zur Überprüfung der Datensicherheit und internen Konsistenz, zur Ausführung von Korrekturen (unter Beteiligung der Autoren) und zur Organisation und Veröffentlichung der geprüften Information.<sup>[11]</sup> Es dürfte keine Überraschung sein, dass die *Strukturberichte* und *Structure Reports* von immensem Wert bei der Aufbereitung des Datenbestands waren.

Allererstes Ziel in der Frühphase war die möglichst rasche Publikation der gesammelten Daten, um Förderinstitutionen zu beruhigen und baldige Akzeptanz in der Fachwelt zu finden. Von 1970 an geschah dies in Form der jährlich erscheinenden *Bibliographic Volumes* innerhalb der *Molecular Structures and Dimensions* (MSD),<sup>[12a]</sup> die in Zusammenarbeit mit der IUCr veröffentlicht wurden. Die Bände waren in 86 chemische „Kapitel“ unterteilt und enthielten eine Vielzahl von Indexen sowie, ab Ausgabe 12, zweidimensionale chemische Diagramme. Die MSD-Serie wurde 1972 um Volume A1 *Interatomic Distances 1960–1965* aufgestockt. Alle MSD-Bände wurden mit der damals neuen Computersatztechnik gedruckt. Ein Gutachter,<sup>[12b]</sup> dem die Größe und das Gewicht der einzelnen Bände aufgefallen war (2,4 kg, 32 × 23 × 5 cm), schlug alternative Gebrauchsmöglichkeiten vor (als Türstopper, Blumenpresse usw.), lobte die Serie aber auch als „maßgeblich für alle Wissenschaftler, die sich mit organischen oder organometallischen Strukturen im kristallinen Zustand befassen. Jegliche Ausflüchte der Unkenntnis der Literatur verlieren nunmehr ihre Gültigkeit.“

Trotz dieser freundlichen Worte wirkten die MSD-Bände, so nützlich sie auch für die manuelle Literatursuche sein mochten, wie ein Nachhall aus der Zeit der gedruckten Kompendien. Um den vollen Wert der CSD nutzen zu können, benötigte man eine spezielle Software für die Datenbanksuche, für die Analyse der Einträge und für die Struktur- und Datenausgabe. Eine erste Software wurde 1978 fertiggestellt<sup>[5a]</sup> und seither ständig weiterentwickelt.<sup>[5b,c]</sup> Der Export des CSD-Systems war aufgrund der vielen verschiedenen Betriebssysteme, die damals auf dem Markt waren, problematisch; Anwender mussten lokale Implementierungen vornehmen. Lochkarten und mehrere hundert Meter lange Magnetbänder waren an der Tagesordnung! Mit dem Aufkommen des VAX-Computers und der Kassettenbänder traten diese Probleme allmählich in den Hintergrund.

#### 4. Die Finanzierungsmodelle des CCDC

Der Aufbau eines großangelegten datenorientierten Systems wie der CSD geschieht nicht zum Nulltarif. Man benötigt erfahrenes Personal mit mannigfaltiger wissenschaftlicher und technischer Expertise, und die technische Infrastruktur muss stetig nachgerüstet werden. Die Finanzierungsmodelle des CCDC lassen sich in zwei Phasen unterteilen. In den späten 1970er Jahren kamen die Fördermittel aus dem britischen Wissenschaftsministerium und dann aus dem Science and Engineering Research Council (SERC), die University of Cambridge stellte Räumlichkeiten und Computeranlagen zur Verfügung. Das Interesse der Universitäten am CSD-System

stieg in den späten 1970er Jahren weltweit rasant. Ein Netzwerk von National Affiliated Centres (NACs) wurde gegründet, die, in Abhängigkeit von der lokalen CSD-Nutzung und auf der Grundlage ökonomischer Faktoren vor Ort, zusätzliche Mittel zur Verfügung stellten. Die NACs waren für den lokalen Betrieb zuständig, einige kleinere Länder schlossen sich zusammen oder wurden direkt von Cambridge beliefert. Ableger dieses System existieren bis heute in Form von aktiven und wertvollen Kollaborationen.

Weiteren Auftrieb erhielt die CSD in den frühen 1980er Jahren, als große Pharmafirmen sich für das System zu interessieren begannen. Modellstudien am Computer – als die Grundlage dessen, was damals als rationale Wirkstoffentwicklung bezeichnet wurde – erforderten intensive Strukturkenntnisse, insbesondere über die bevorzugten Konformationen von Molekülen und die intermolekularen Wechselwirkungen der chemischen Gruppen. Ebenso brauchte man Daten, um Kraftfelder zu parametrisieren, oder als experimentellen Ausgangspunkt für die automatisierte Generierung von Konformeren und für Protein-Ligand-Docking. Die in der CSD enthaltenen Daten waren ideal für solche Anwendungen und sind bis heute ein integraler Bestandteil der modernen Wirkstoffforschung.<sup>[13]</sup> Auch andere Sparten wie die Feinchemikalienindustrie nutzten das CSD-System.

Die Realisierung finanzieller Erlöse führte zu Diskussionen mit dem SERC und der Universität, und 1987 begann das CCDC, sich vom Bezug öffentlicher Fördergelder zu lösen. Das CCDC wurde 1989 vollständig unabhängig und firmiert seither als eine eigenfinanzierte gemeinnützige Einrichtung. Das neue CCDC blieb eng mit der Universität verbunden und ist heute ein Partnerinstitut der Universität, an dem auch Doktoranden ausgebildet werden. Die Geschäfte des CCDC werden von einem internationalen Kuratorium beaufsichtigt, das die Interessen der Wissenschaftsgemeinde vertritt und strategisch und wissenschaftlich berät.

#### 5. Das CSD-System der Gegenwart

Die CSD enthält heute Informationen zu nahezu 700 000 Kristallstrukturen, der jährliche Anstieg ist in Abbildung 1

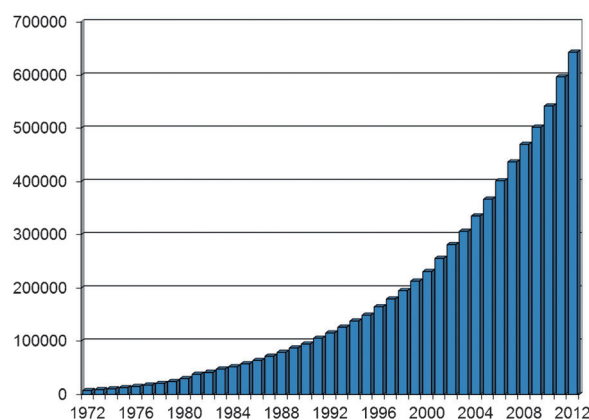


Abbildung 1. Jährliches Wachstum der Cambridge Structural Database von 1970 bis 2012.

dargestellt. Daten von ungefähr 250 000 Strukturen wurden manuell erfasst, und alle Daten sind sorgfältig auf Tippfehler und Konsistenz geprüft. Ein Wendepunkt bei der Datenerfassung war Anfang der 1990er Jahre die Entwicklung des Crystallographic Information File (CIF)<sup>[14]</sup> (in Zusammenarbeit mit der IUCr) als ein allgemeines Format für die elektronische Übermittlung von Kristallstrukturdaten. Die IUCr<sup>[15]</sup> und das CCDC<sup>[16]</sup> stellen heute Software für die Überprüfung von CIF-Dateien zur Verfügung, und die Häufigkeit von Fehlern ist drastisch gesunken. Anfang 2013 wurde eine neue interne Software installiert, die die Konvertierung der CIF-Rohdaten in die endgültigen CSD-Einträge beschleunigt und eine Reihe von altmodischen Büroarbeiten überflüssig machte. Diese Software beinhaltet wissenschaftliche Module für die automatisierte Datenauswertung, z. B. deCIFer.<sup>[17]</sup> Ein kurzer statistischer Überblick über den Informationsgehalt der CSD ist in Tabelle 1 gegeben.

Die Anwendersoftware wurde in den letzten 15 Jahren ebenfalls deutlich weiterentwickelt. *ConQuest*<sup>[18a]</sup> führt Suchen in allen Informationsfeldern des CSD durch und kann eine zweidimensionale Substruktursuche mit dreidimensionalen geometrischen Randbedingungen kombinieren, um Wasserstoffbrücken oder andere nichtbindende Wechselwirkungen aufzuspüren. Eine Onlineversion, WebCSD, ist ebenfalls verfügbar (Abbildung 2).<sup>[18b]</sup> *Mercury*,<sup>[18a,20]</sup> das Strukturzeichenprogramm der CCDC, wurde zu einer umfangreichen Analysesoftware für Strukturen und geometrische Daten weiterentwickelt. Neben den Standardoptionen kann *Mercury* (Abbildung 3) intermolekulare Wechselwirkungen, H-verbrückte Synthone,<sup>[23]</sup> ausgedehnte Netzwerke und Graphensatz-Deskriptoren<sup>[24]</sup> sowie computerberechnete Pulverbeugungsmuster darstellen. *Mercury* enthält außerdem Module für die Analyse von mittels *ConQuest* gewonnener geometrischer Daten.<sup>[25]</sup>

**Tabelle 1:** Zahlen zur Cambridge Structural Database, Stand 1. Juli 2013.

		[%] der CSD
Strukturen gesamt	658 059	100.0
unterschiedliche Verbindungen	601 308	–
Zahl der Literaturquellen	1518	–
organische Strukturen	280 809	42.6
mit Übergangsmetallen	353 201	53.7
mit Li–Fr oder Be–Ra	33 011	5.0
mit Hauptgruppenmetallen	40 166	6.1
Strukturen mit 3D-Koordinaten	614 824	93.4
fehlerfreie Koordinaten	604 539	98.3 <sup>[a]</sup>
Neutronenstudien	1583	0.2
Pulverbeugungsstudien	2721	0.4
Tief-/Hochtemperaturstudien	288 213	43.9
absolute Konfiguration bestimmt	13 510	2.1
Strukturen mit Fehlordnung	149 994	22.8
polymorphe Strukturen	19 990	3.0
R-Faktor < 0.100	618 027	93.9
R-Faktor < 0.075	560 089	85.1
R-Faktor < 0.050	361 367	54.9
R-Faktor < 0.030	74 501	11.3
Atome mit 3D-Koordinaten	50 821 771	–

[a] Prozent der Strukturen, für die Daten in der CSD hinterlegt sind.

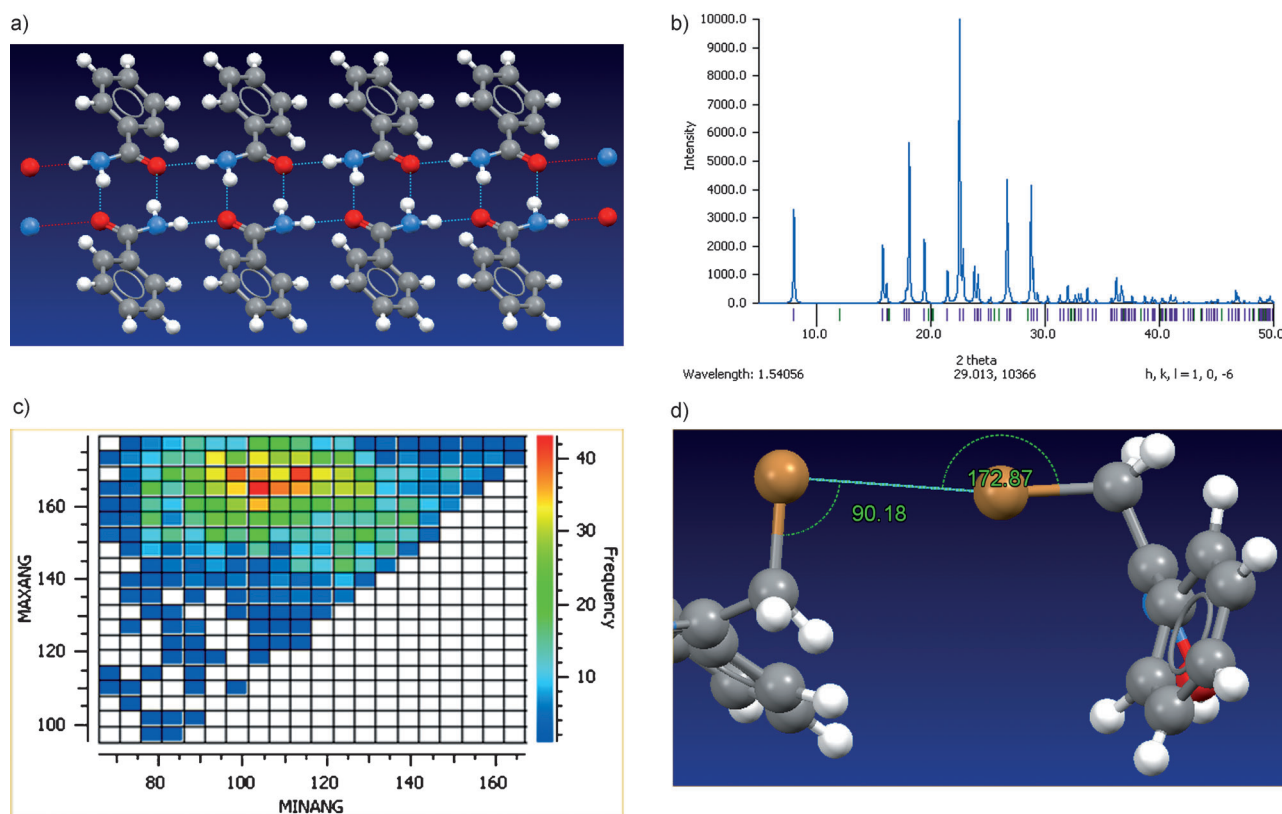
Das CSD-System beinhaltet zwei ausführliche Informationsdatenbanken. *Mogul*<sup>[26a]</sup> enthält mehr als 20 Millionen Bindungslängen, Bindungswinkel und Drehwinkel, die in > 1.5 Millionen chemisch suchbare geometrische Verteilungen, jeweils entsprechend einer spezifischen chemischen Umgebung gegliedert sind (Abbildung 4). Eine jüngste Erweiterung<sup>[26b]</sup> ermöglicht außerdem die Suche nach Konformationsdaten chemischer Ringe. *Mogul* kann auch die Drehwinkelverteilungen für alle frei drehbaren Bindungen eines eingegebenen Moleküls erzeugen oder alle geometrischen Parameter gegen Mittelwerte aus der CSD abgleichen,

The screenshot shows the WebCSD interface for the entry CAFINE (1,3,7-Trimethyl-purine-2,6-dione monohydrate). The left sidebar contains a search bar and a list of entries. The main area displays a 3D ball-and-stick model of the molecule. The right panel shows a detailed information table for the entry.

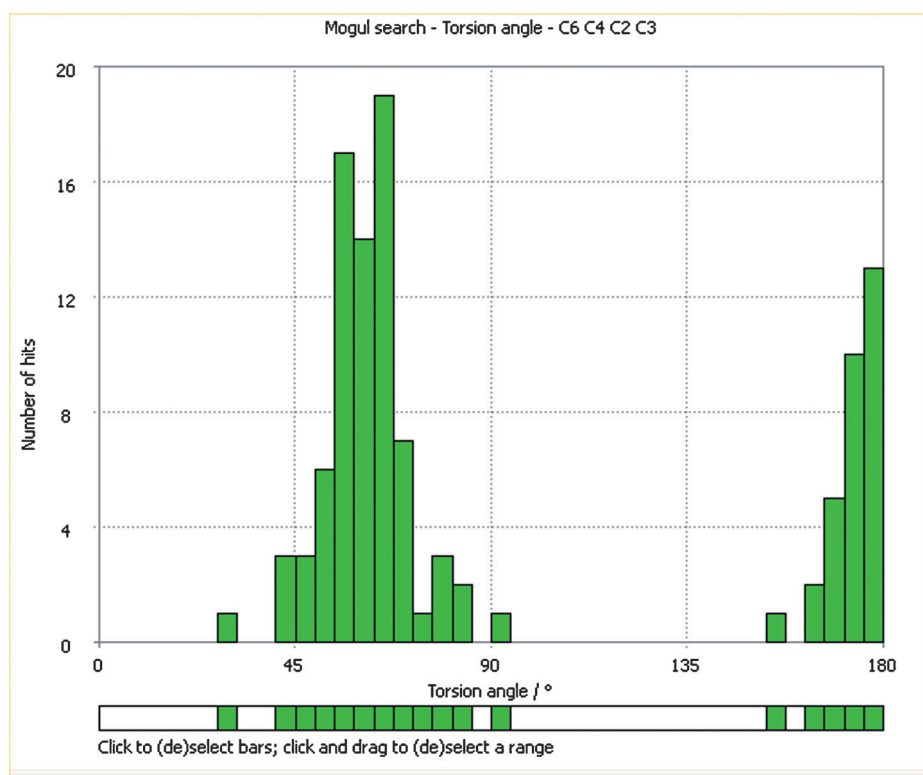
Field	Value
Identifier	CAFINE
Author(s)	D. J. Sutor
Reference	Acta Crystallogr. (1958), 11, 453, doi:10.1107/S0365110X58001286
Formula	C <sub>8</sub> H <sub>10</sub> N <sub>4</sub> O <sub>2</sub> ·H <sub>2</sub> O
Compound	1,3,7-Trimethyl-purine-2,6-dione monohydrate
SMILES	CN1C=NC2=C1C(=O)N(C)C(=O)N2C.O
Synonym	Caffeine monohydrate
Space Group	P 2 <sub>1</sub> /a
Cell Lengths	a 14.8(1) b 16.7(1) c 3.97(3)
Cell Angles	α 90 β 97.0(5) γ 90
Cell Volume	973.911
Z, Z'	Z: 4 Z': 1
R-Factor (%)	14.6
Reduced Cell Lengths	a 3.97 b 14.8 c 16.7
Reduced Cell Angles	α 90 β 90 γ 97
Reduced Cell Volume	973.911
Temperature (K)	Room Temp. (283-303)
Density (CCDC)	1.447
Average Sigma (C-C)	0.01 < sigma <= 0.03
Radiation Probe	x-ray
Bioactivity	stimulant agent which increases CNS activity

**Abbildung 2.** Informationstafel für Caffein-Monohydrat (CSD-Code CAFINE<sup>[19]</sup>) in WebCSD.

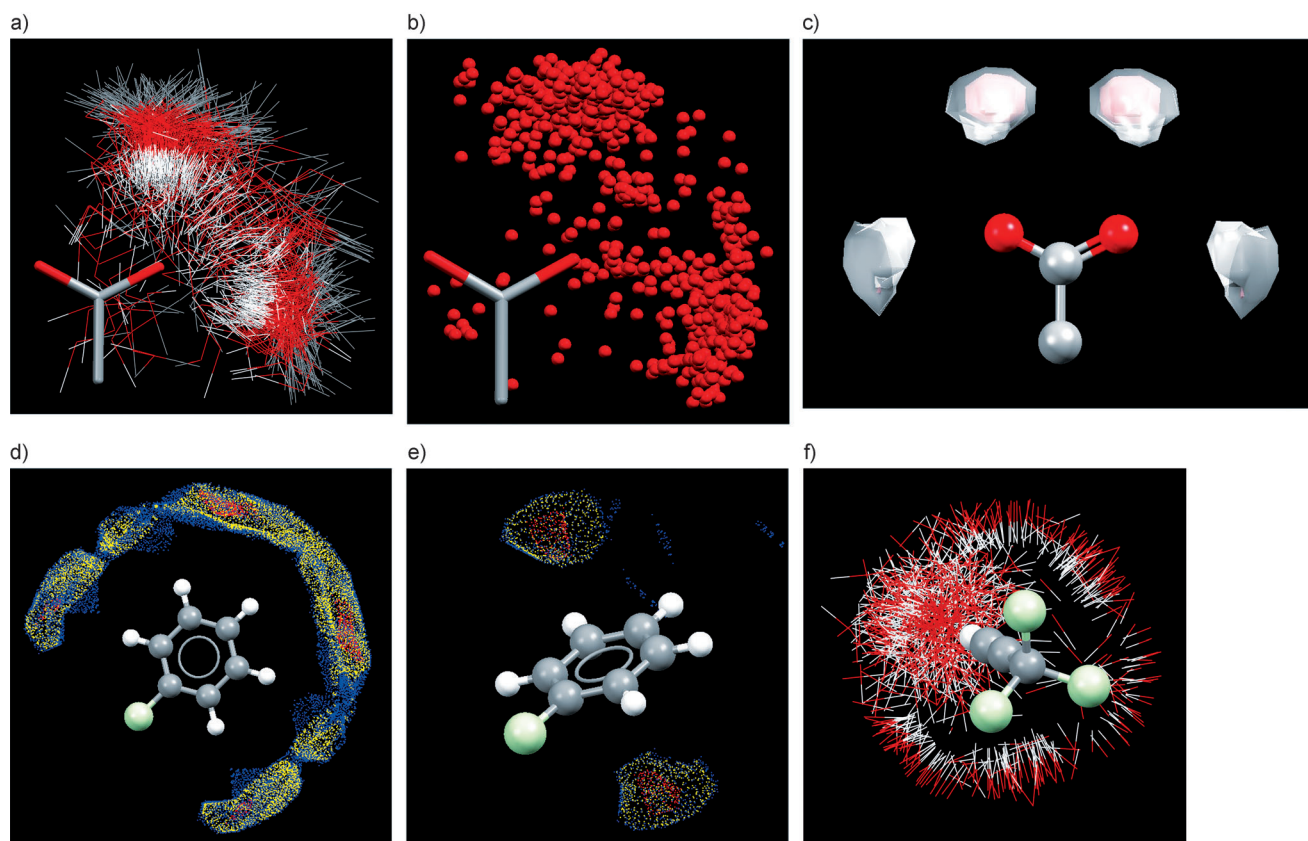




**Abbildung 3.** Beispiele für Strukturdarstellungen mit Mercury: a) Wasserstoffbrückenstruktur von Benzamid (CSD-Code BZAMID<sup>[21]</sup>), b) berechnetes Pulverbeugungsmuster für BZAMID,<sup>[21]</sup> c) Farbtonkarte der beiden C-Hal-Hal-Winkel bei Halogen-Halogen-Wechselwirkungen, die bevorzugte Winkel von ungefähr 90° bzw. 180° anzeigt, entsprechend der Struktur in (d) (CSD-Code ABACOX10).<sup>[22]</sup>



**Abbildung 4.** Verteilung der C-C-C-C-Drehwinkel in 2,3-Dimethylbutan-Fragmenten ( $C_{sp^3}$ )<sub>2</sub>-CH-CH-( $C_{sp^3}$ )<sub>2</sub>, ermittelt aus der Mogul-Informationdatenbank.



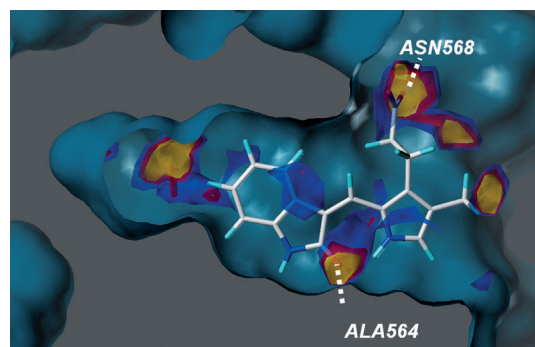
**Abbildung 5.** Die IsoStar-Informationsdatenbank. Räumliche Verteilung von O-H-Donoren um eines der äquivalenten Sauerstoffatome einer geladenen Carboxylatgruppe unter Verwendung von a) CSD-Daten und b) PDB-Daten; c) Konturdiagramm der vollständig symmetrisierten Verteilung um das Carboxylat. Verteilung von d) Carbonyl-O-Atomen, e) von aliphatischen C-H-Atomen um einen Phenylring und f) von O-H-Gruppen um eine Ethinylgruppe  $C_{sp}-C\equiv C-H$ . Man beachte in (f) die Bildung von  $C\equiv CH\cdots O$ -Bindungen entlang der Richtung der Ethinylbindung und die Bildung eines Rings von  $OH\cdots\pi(C\equiv C)$ -Bindungen senkrecht dazu.

eine Funktion, die für die Strukturlösung und Verfeinerung von neuartigen Kristallstrukturen sowohl niedermolekularer Verbindungen<sup>[27]</sup> als auch proteingebundener Liganden nützlich ist.<sup>[28]</sup> IsoStar<sup>[29]</sup> ist eine Bibliothek mit graphischen und numerischen Daten nichtbindender Wechselwirkungen, die interaktive 3D-Streudiagramme liefert, in denen die Verteilung einer von 48 Kontaktgruppen, z. B. einem H-Brücken-Donor, um eine zentrale Gruppe visualisiert ist (Abbildung 5). Die 300 zentralen Gruppen decken einen breiten Bereich von chemischen Funktionalitäten ab. Mehr als 25 500 Streudiagramme sind verfügbar, davon sind > 20 000 aus der CSD und > 5500 von Protein-Ligand-Komplexen in der PDB abgeleitet. Ungefähr 1500 ab initio berechnete Potentialenergieminima<sup>[30]</sup> sind ebenfalls in IsoStar enthalten.

## 6. Diversifikation der CCDC-Software

Um eine stärkere Vernetzung mit der pharmazeutischen und kristallographischen Forschung zu erreichen, hat das CCDC seine Benutzersoftware diversifiziert und an spezielle Anforderungen angepasst. Zu dieser Software gehört nun GOLD<sup>[31]</sup> ein Programm für Protein-Ligand-Docking, das Konformationen und Wasserstoffbrückendaten aus der CSD zur Parametrisierung und in Scoring-Funktionen nutzt; Su-

perStar<sup>[32]</sup> das Wechselwirkungszentren in Proteinbindetaschen auf der Grundlage von experimentellen Daten aus IsoStar kartiert (Abbildung 6); Relibase<sup>[33]</sup> ein aus der PDB<sup>[6c,d]</sup> abgeleitetes Datenbanksystem, das eine vielseitige Suche nach Proteinen, Liganden und ihren Wechselwirkun-



**Abbildung 6.** Mittels SuperStar ermittelte Wechselwirkungszentren („Hot-Spots“, goldfarben) für eine C=O-Gruppe in der Bindungsregion der Tyrosinkinase; die experimentell bestimmte Position eines Oxindol-Inhibitors ist als Überlagerung gezeigt. Die Übereinstimmung zwischen der experimentellen Bindestellen der C=O-Gruppen und den mittels SuperStar ermittelten Hot-Spots ist eindeutig erkennbar.

gen erlaubt; und *DASH*,<sup>[34]</sup> das direkte Methoden für die Strukturlösung aus Röntgenpulverdaten nutzt und mit *Mogul*<sup>[26]</sup> direkt verlinkt ist, um Konformationsdaten zu liefern<sup>[35]</sup> und auf diese Weise den Suchraum zu reduzieren.

## 7. Die CSD als Katalysator wissenschaftlicher Forschung

Von den frühesten Tagen an wurden CSD-Daten routinemäßig für Vergleichszwecke in Kristallstrukturberichten zitiert. Von grundlegender Bedeutung ist, dass die CSD als Stützpfeiler für die Forschung in Disziplinen diene, in denen die Kenntnis chemischer Strukturen maßgeblich ist. Rund 3000 Veröffentlichungen dieser Art sind bisher in der Literatur erschienen<sup>[36a]</sup> und waren jüngst Gegenstand einer Zitationsanalyse.<sup>[36b]</sup>

Etliche frühe CSD-basierte Studien betrafen die Analyse von Reaktionspfaden und Strukturkorrelationen.<sup>[37a]</sup> Die erste Studie dieser Art war eine Arbeit von Bürgi und Dunitz,<sup>[37b]</sup> die den Reaktionspfad für den Angriff eines Stickstoff-Nukleophils auf ein Carbonylzentrüm kartierten. Die CSD-basierte Forschung breitete sich rasch auf andere Anwendungen aus, einschließlich systematischer Studien intra- und intermolekularer Systeme.

Eine frühe Veröffentlichung belegte die Fähigkeit von C-H-Donoren, Wasserstoffbrücken mit O-, N- und Cl-Akzeptoren zu bilden.<sup>[38]</sup> Auf diese Veröffentlichung folgte eine Flut von CSD-basierten Studien intermolekularer Wechselwirkungen – starker und schwacher, konventioneller und unkonventioneller –,<sup>[39]</sup> die alle zur Formulierung von supramolekularen Synthons<sup>[23]</sup> und Graphensatz-Deskriptoren<sup>[24]</sup> von Wasserstoffbrückennetzwerken beitrugen. Ebenfalls in der Frühphase wurden die vorhandenen Tabellen von Bindungslängen<sup>[8]</sup> durch definitive Standards ersetzt,<sup>[40]</sup> und die Forschung auf intramolekularer Ebene weitete sich auf Substituenteneffekte, detaillierte Studien von cyclischen und acyclischen Konformationen und Studien der Koordinationsumgebung von Metallzentren aus. CSD-basierte Forschung leistet nach wie vor einen signifikanten Beitrag in der organischen und metallorganischen Chemie, beim Kristall-Engineering, in der Kristallstrukturvorhersage, für die Analyse von Protein-Ligand-Wechselwirkungen, in der Wirkstoffentwicklung und in den Materialwissenschaften; diese Beiträge wurde an mehreren Stellen zusammengefasst.<sup>[13,39,41]</sup> Die American Chemical Society gab jüngst bekannt,<sup>[42]</sup> dass der derzeitige Standardliteraturverweis auf das CSD-System<sup>[5c]</sup> die am häufigsten zitierte Veröffentlichung aus der Chemie seit 2002 ist.

## 8. Das CCDC als wissenschaftliches Institut

Die Forschung am CCDC nahm im Lauf der Jahre eine natürliche Entwicklung von reinen Kristallstrukturbestimmungen hin zu gezielten Anwendungen der Datenbank, und mittlerweile existieren fast 800 Veröffentlichungen von Autoren des CCDC. Einige dieser Veröffentlichungen beschreiben Entwicklungen am CSD-System, z. B. die Einführung von

*Mogul* und *IsoStar* und die Validierung von *SuperStar* und *GOLD*. Andere resultieren aus Arbeiten von Doktoranden und Gastwissenschaftlern, der Hauptanteil der Forschung am CCDC wird aber von festen Mitarbeitern betrieben. Es existiert eine enge Kooperation mit Industriepartnern, vor allem im Bereich der Wirkstoffentwicklung und -formulierung.<sup>[43]</sup>

Etwa 90 % der Formulierungen niedermolekularer Wirkstoffe sind kristallin. Ihre übliche Verabreichungsform ist als Salze oder Kokristalle mit Trägerstoffen. Mit der neu entwickelten Software *Mercury* steht Forschern nun ein Werkzeug zur Verfügung, um die Bildung und Stabilität von kristallinen Festkörpern besser zu verstehen. *Mercury* kann nach supramolekularen Synthons aus spezifizierten funktionellen Gruppen suchen, ebenso nach Packungsmustern, wobei der Nutzer ein bestimmtes Muster aus einer vollständigen CSD-Struktur selektieren kann; auch eine Suche nach Ähnlichkeiten in Kristallpackungen ist möglich.<sup>[20]</sup>

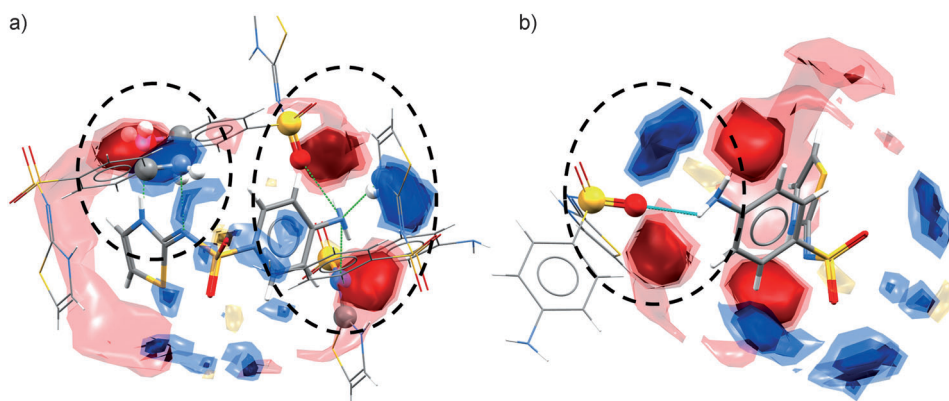
Polymorphie kann ein kritisches Problem für die Arzneimittelformulierung sein. Ergänzend zum experimentellen Polymorph-Screening kann mittels CSD eine datenbankbasierte Propensitätsanalyse von Wasserstoffbrücken ausgeführt werden.<sup>[44]</sup> Ausgehend von einem einfachen zweidimensionalen chemischen Diagramm einer Zielstruktur werden CSD-Daten von verwandten Molekülen verwendet, um die Wahrscheinlichkeit der Bildung jeder potenziellen H-Brücke in der Kristallstruktur des Moleküls zu berechnen. Hoch wahrscheinliche oder unwahrscheinliche H-Brücken werden rasch aufgedeckt und auf diese Weise mögliche Stabilitätsprobleme aufzufindig gemacht, wie am Beispiel einer Studie zu Ritonavir gezeigt wurde.<sup>[45]</sup> Solche Analysen können auch genutzt werden, um Mehrkomponentenkristalle zu untersuchen, z. B. Kokristalle oder Solvate, bei denen die zweite Komponente oft alternative Möglichkeiten für Donor-Akzeptor-Wechselwirkungen einführt.<sup>[46]</sup>

Derzeitige Ansätze für die Erforschung von Mehrkomponentensystemen nutzen auch einige der am CCDC entwickelten Werkzeuge für die Wirkstoffentdeckung. Ein kürzlich entwickeltes Verfahren<sup>[47]</sup> behandelt supramolekulares Design als ein Docking-Problem analog zum Protein-Ligand-Docking. Hierbei wird die bekannte Struktur eines reinen Wirkstoffs als Wirkkomponente vorgegeben und mittels der *GOLD*-Methodik mit potenziellen Kokristallbildnern kombiniert. In einem Verfahren<sup>[48]</sup> werden *IsoStar*,<sup>[29]</sup> *SuperStar*<sup>[32]</sup> und *Mercury*<sup>[20]</sup> verwendet, um vollständige Wechselwirkungskarten von Wirkstoffmolekülen mit funktionellen Gruppen zu erzeugen. Diese Methode wird für Stabilitätsstudien von Polymorphen eingesetzt, wie in Abbildung 7 gezeigt.

## 9. Die Zukunft

Die CSD bildet das Zentrum unserer Aktivitäten, und das CCDC ist gut aufgestellt, um der sicher auch künftig ansteigenden Produktivität bei der Ermittlung von Kristallstrukturen gerecht zu werden. Neue Datenverarbeitungssysteme eröffnen neue Horizonte. Durch den Wegfall historisch bedingter Formatbeschränkungen kann die CSD nun atomare





**Abbildung 7.** Wechselwirkungskarten für Sulfathiazol in den Packungsdiagrammen von a) Polymorph V (SUTHAZ19<sup>[49]</sup>) und b) Polymorph I (SUTHAZ16<sup>[49]</sup>). Bevorzugte Akzeptorpositionen sind rot markiert, Donorpositionen sind blau. In (a) befindet sich für jedes der Wechselwirkungsmaxima ein passender Donor oder Akzeptor innerhalb der Konturfläche, während in (b) einer der Akzeptoren außerhalb der berechneten Wechselwirkung liegt. Ein „unzufriedener“ Donor oder Akzeptor ist ein Indiz für Metastabilität, wie hier bei Polymorph I der Fall. Einzelheiten siehe Lit. [48].

Auslenkungsparameter, verbesserte Beschreibungen von Fehlern, Metalloxidationszustände und zusätzliche Datenelemente, wie sie aus den hinterlegten CIFs routinemäßig verfügbar sind, verarbeiten.

Eines der wichtigsten Themen für die Zukunft betrifft die Forderung nach kostenlosem Zugang (Open Access; OA) zu Kristallstrukturdaten. In den letzten 15 bis 20 Jahren sind die CIF-Rohdaten von Kristallstrukturstudien immer mehr frei verfügbar geworden, entweder über die betreffende Zeitschrift oder über den kostenlosen CIF-Service („request a structure“).<sup>[50]</sup> Dies führte zur Neugründung zweier CIF-Datenbanken, der Crystallography Open Database (COD),<sup>[51a,b]</sup> die hinterlegte und per Download bezogene CIFs enthält, und CrystalEye<sup>[51c,d]</sup> mit automatisch aus dem Internet gezogenen CIFs (obwohl diese Datendank anscheinend seit Mitte 2011 kein Update mehr erhalten hat). Beide Datenbanken haben von verschiedener Seite externe Fördermittel erhalten (oftmals zeitlich befristet) und sind also nicht im absoluten Sinne „kostenlos“. In allen Open-Access-Modellen gibt es jemanden, der zahlt, für gewöhnlich ein öffentlicher Träger, eine Institution oder die Autoren selbst, die auf diese Weise dem Leser oder Nutzer den kostenlosen Zugriff ermöglichen.

Die in diesen Datenbanken enthaltenen CIFs sind natürlich auch Bestandteil der CSD, neben vielen anderen Informationen. Zu bedenken ist allerdings, dass die Bereitstellung von Daten in einer vollständig retrospektive, lückenlose und wissenschaftlich kuratierte Datenbank mit qualitativ hochwertiger Zugriffssoftware und Nutzersupport eine finanziell stabile und dauerhafte Organisation erfordert. Eine Förderung durch eine nationale Regierung (oder einen Zusammenschluss von Regierungen wie in der EU) ist dem Risiko gesamtwirtschaftlicher Krisen und politischer Unwägbarkeiten der Mittelvergabe ausgesetzt. Diese Risiken sind real. Ein breit angelegtes Förderkonzept beseitigt das Risiko nicht gänzlich, mindert es aber beträchtlich. Das Finanzierungsmodell des CCDC, das international angelegt ist und auf den Beiträgen der Datenbank- und Softwarenutzer an Universitäten und aus der Industrie gründet, dürfte die solideste Lösung für die dauerhafte Instandhaltung einer spezialisierten

wissenschaftlichen Bezugsquelle sein. Der Verweis auf „dauerhaft“ ist hier wichtig: Schon ein kurzzeitiger Ausfall von Fördergeldern kann die Kernaufgaben eines Datenzentrums wie des CCDC schwerwiegend beeinträchtigen. Diese Aufgaben umfassen die Pflege eines Datenbanksystems, das so vollständig und genau wie möglich sein muss, um die wissenschaftliche Integrität der historischen Aufzeichnungen zu gewährleisten und ihren Wert für die Wissenschaftsgemeinde zu maximieren.<sup>[52]</sup> Dies mögen hohe und vielleicht auch unerreichbare Ideale sein, aber wir sollten zumindest versuchen, sie zu erreichen.

## 10. Schlussbemerkungen

Wir schließen mit einem anderen Jubiläum: 2015 feiert das CCDC das 50. Jahr seiner Gründung, die fast genau ein halbes Jahrhundert nach der Bestimmung der ersten Kristallstruktur erfolgte.<sup>[1]</sup> Während dieser 50 Jahre bewältigte das CCDC all die wissenschaftlichen, technischen und finanziellen Aspekte, die 1964 in der ursprünglichen Ausschreibung formuliert worden waren. Die Fokussierung auf niedermolekulare Strukturen erwies sich dabei als mehr als gerechtfertigt: Wie schon 1929 realisiert wurde, ist die Kristallographie in der Tat eine besondere Analysetechnik, und der Wert kristallographischer Daten für die Wissenschaftsgemeinde ist seit den 1960er Jahren mit jedem Jahrzehnt weiter gestiegen. Die Ausschreibung war nicht nur wissenschaftlich visionär, sie kam auch gerade zum richtigen Zeitpunkt: Hätte das CCDC nicht in einem relativ kleinen Zeitfenster in der Mitte der 1960er Jahre seinen Betrieb aufgenommen, wäre es möglicherweise überhaupt nie gegründet worden. Die Zahl der veröffentlichten Kristallstrukturen hatte eine Größenordnung erreicht, die eine Datenbank zu einem wertvollen Werkzeug machte, gleichzeitig war der Rückbestand an Strukturen noch nicht unüberwindlich. Betrachtet man den rapiden Anstieg an Strukturbestimmungen, hätte dieser Fall jedoch sehr bald eintreten können (siehe Abbildung 2). Jedes Jahr der Verzögerung hätte ein erheblich gestiegenes Anfangspensum und erheblich höhere Kosten bedeutet.<sup>[52]</sup>



Vor allem anderen hat die CSD in höchstem Maß von der Unterstützung der kristallographischen Fachwelt profitiert – die Koordinaten jeder organischen und metallorganischen Verbindung, für die eine Kristallstruktur veröffentlicht wurde, in einer einzelnen Datenbank verfügbar zu haben, ist etwas, auf das die Disziplin zurecht stolz ist.

Eingegangen am 24. Juli 2013

- [1] a) W. L. Bragg, *Proc. Cambridge Philos. Soc.* **1912**, 17, 43–57; b) W. L. Bragg, *Proc. R. Soc. London Ser. A* **1913**, 89, 248–277; c) W. H. Bragg, *Proc. R. Soc. London Ser. A* **1913**, 89, 430–438.
- [2] a) P. P. Ewald, C. Hermann, *Strukturbericht 1913–1928*, Akademische Verlagsgesellschaft, Leipzig, **1929**; b) *Structure Reports 1940–1950*, Vol. 8–13 (Hrsg.: A. J. C. Wilson, N. C. Baenziger, J. M. Bijvoet, J. M. Robertson), Reidel, Dordrecht, **1990**. Siehe <http://www.iucr.org/books/structure-reports>.
- [3] L. Pauling, *J. Am. Chem. Soc.* **1929**, 51, 1010–1026.
- [4] <http://www.iucr.org/people/nobel-prize>.
- [5] a) F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, D. G. Watson, *Acta Crystallogr. Sect. B* **1979**, 35, 2332–2339; b) F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith, D. G. Watson, *J. Chem. Inf. Comput. Sci.* **1991**, 31, 187–204; c) F. H. Allen, *Acta Crystallogr. Sect. B* **2002**, 58, 380–388.
- [6] a) ICSD: A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, *Acta Crystallogr. Sect. B* **2002**, 58, 364–369; b) CRYSTMET: P. S. White, J. R. Rodgers, Y. Le Page, *Acta Crystallogr. Sect. B* **2002**, 58, 343–348; c) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, 28, 235–242; d) <http://www.wwpdb.org/>; e) NDB: H. M. Berman, J. Westbrook, Z. Feng, L. Iype, B. Schneider, C. Zardecki, *Acta Crystallogr. Sect. D* **2002**, 58, 889–898.
- [7] J. D. H. Donnay, G. Donnay, E. G. Cox, O. Kennard, M. V. King, *Crystal Data: Determinative Tables*, American Crystallographic Association, **1963**.
- [8] L. E. Sutton, *Tables of Interatomic Distances in Molecules and Ions*, The Chemical Society, London, Special Publications Nos. 11 and 18, **1959** und **1965**.
- [9] a) The Royal Society Scientific Information Conference: report and papers submitted, The Royal Society, London, **1948**; b) siehe auch: B. Vickery, *J. Doc.* **1998**, 54, 281–283.
- [10] O. Kennard, *In the Beginning was Bernal*, **1988** (persönliche Mitteilung von Vorlesungsschriften).
- [11] a) F. H. Allen, O. Kennard, W. D. S. Motherwell, W. G. Town, D. G. Watson, T. J. Scott, A. C. Larson, *J. Appl. Crystallogr.* **1974**, 7, 73–78; b) F. H. Allen, O. Kennard, D. Watson, K. M. Crennell, *J. Chem. Inf. Comput. Sci.* **1982**, 22, 129–139; c) F. H. Allen, *J. Chem. Inf. Comput. Sci.* **1980**, 20, 68–76.
- [12] a) *Molecular Structures and Dimensions, Bibliographic Vol. 1–16 and Vol. A1* (Hrsg.: O. Kennard, D. G. Watson und andere), Reidel, Dordrecht, **1970–1986**; b) G. A. Jeffrey, *Acta Crystallogr. Sect. B* **1978**, 34, 3847.
- [13] a) K. A. Brameld, B. Kuhn, D. C. Reuter, M. Stahl, *J. Chem. Inf. Model.* **2008**, 48, 1–24; b) C. Bissantz, B. Kuhn, M. Stahl, *J. Med. Chem.* **2010**, 53, 5061–5084; c) B. Kuhn, P. Mohr, M. Stahl, *J. Med. Chem.* **2010**, 53, 2601–2611.
- [14] S. R. Hall, F. H. Allen, I. D. Brown, *Acta Crystallogr. Sect. A* **1991**, 47, 655–685.
- [15] <http://journals.iucr.org/e/services/authorservices.html>.
- [16] F. H. Allen, O. Johnson, G. P. Shields, B. R. Smith, M. Towler, *J. Appl. Crystallogr.* **2004**, 37, 335–338.
- [17] I. J. Bruno, G. P. Shields, R. Taylor, *Acta Crystallogr. Sect. B* **2011**, 67, 333–349.
- [18] a) I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, R. Taylor, *Acta Crystallogr. Sect. B* **2002**, 58, 389–397; b) I. R. Thomas, I. J. Bruno, J. C. Cole, C. F. Macrae, E. Pidcock, P. A. Wood, *J. Appl. Crystallogr.* **2010**, 43, 362–366.
- [19] D. J. Sutor, *Acta Crystallogr.* **1958**, 11, 453–458.
- [20] a) C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler, J. van de Streek, *J. Appl. Crystallogr.* **2006**, 39, 453–457; b) C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, P. A. Wood, *J. Appl. Crystallogr.* **2008**, 41, 466–470.
- [21] B. R. Penfold, J. C. B. White, *Acta Crystallogr.* **1959**, 12, 130–135.
- [22] J. B. Wetherington, J. W. Moncrief, *Acta Crystallogr.* **1973**, 29, 1520–1525.
- [23] a) G. R. Desiraju, *Angew. Chem.* **1995**, 107, 2541–2558; *Angew. Chem. Int. Ed. Engl.* **1995**, 34, 2311–2327; b) G. R. Desiraju, *J. Am. Chem. Soc.* **2013**, 135, 9952–9967.
- [24] J. Bernstein, R. E. Davis, L. Shimoni, N.-L. Chang, *Angew. Chem.* **1995**, 107, 1689–1708; *Angew. Chem. Int. Ed. Engl.* **1995**, 34, 1555–1573.
- [25] R. A. Sykes, P. McCabe, F. H. Allen, G. M. Battle, I. J. Bruno, P. A. Wood, *J. Appl. Crystallogr.* **2011**, 44, 882–886.
- [26] a) I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris, A. G. Orpen, *J. Chem. Inf. Comput. Sci.* **2004**, 44, 2133–2144; b) S. J. Cottrell, T. S. G. Olsson, R. Taylor, J. C. Cole, J. W. Liebeschuetz, *J. Chem. Inf. Model.* **2012**, 52, 956–962.
- [27] D. J. Watkin, R. J. Cooper, <http://www.xtl.ox.ac.uk/crystals.html>.
- [28] J. W. Liebeschuetz, J. Hennemann, T. S. G. Olsson, C. R. Groom, *J. Comput.-Aided Mol. Des.* **2012**, 26, 169–183.
- [29] I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk, *J. Comput.-Aided Mol. Des.* **1997**, 11, 525–537.
- [30] I. C. Hayes, A. J. Stone, *J. Mol. Phys.* **1984**, 53, 83–105.
- [31] a) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, 267, 727–748; b) M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins* **2003**, 52, 609–623; c) J. W. Liebeschuetz, J. C. Cole, O. Korb, *J. Comput.-Aided Mol. Des.* **2012**, 26, 737–748.
- [32] a) M. L. Verdonk, J. C. Cole, R. Taylor, *J. Mol. Biol.* **1999**, 289, 1093–1108; b) M. L. Verdonk, J. C. Cole, P. Watson, V. Gillet, P. Willett, *J. Mol. Biol.* **2001**, 307, 841–859; c) J. W. M. Nissink, C. W. Murray, M. J. Hartshorn, M. L. Verdonk, J. C. Cole, R. Taylor, *Proteins* **2002**, 49, 457–471.
- [33] a) M. Hendlich, A. Bergner, J. Guenther, G. Klebe, *J. Mol. Biol.* **2003**, 326, 607–620; b) J. Guenther, A. Bergner, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2003**, 326, 621–636.
- [34] W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell, J. Cole, *J. Appl. Crystallogr.* **2006**, 39, 910–915.
- [35] E. Pidcock, J. van de Streek, M. U. Schmidt, *Z. Kristallogr.* **2007**, 222, 713–717.
- [36] a) <http://www.ccdc.cam.ac.uk/ResearchAndConsultancy/CCDCResearch/Pages/WebCite.aspx>; b) R. Wong, F. H. Allen, P. Willett, *J. Appl. Crystallogr.* **2010**, 43, 811–824.
- [37] a) H.-B. Bürgi, J. D. Dunitz, *Structure Correlation*, VCH, Weinheim, **1997**; b) H.-B. Bürgi, J. D. Dunitz, *Acc. Chem. Res.* **1983**, 16, 153–161.
- [38] a) R. Taylor, O. Kennard, *Acc. Chem. Res.* **1984**, 17, 320–326. Faszinierende historische Übersichten über die Kontroversen, die es vor der Veröffentlichung dieser Arbeit gegeben hat, finden sich bei: b) C. Schwalbe, *Cryst. Rev.* **2012**, 18, 191–206 und c) J. Bernstein, *Cryst. Growth Des.* **2013**, 13, 961–964.

- [39] Siehe z.B.: a) G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer, Berlin, **1991**; b) G. R. Desiraju, T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, Oxford, **1999**; c) M. Nishio, *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873–13900.
- [40] a) F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *J. Chem. Soc. Perkin Trans. 2* **1987**, S1–S19; b) A. G. Orpen, L. Brammer, F. H. Allen, O. Kennard, D. G. Watson, R. Taylor, *J. Chem. Soc. Dalton Trans.* **1989**, S1–S83.
- [41] a) F. H. Allen, W. D. S. Motherwell, *Acta Crystallogr. Sect. B* **2002**, *58*, 407–422; b) R. Taylor, *Acta Crystallogr. Sect. D* **2002**, *58*, 879–888; c) F. H. Allen, J. A. Chisholm, P. A. Wood, P. T. A. Galek, L. Fábíán, O. Korb, A. J. Cruz-Cabeza, J. W. Liebeschuetz, C. R. Groom, E. Pidcock, *Supramolecular Chemistry: From Molecules to Nanomaterials*, Wiley, Chichester, **2012**, S. 2927–2946.
- [42] E. K. Wilson, *Chem. Eng. News* **2012**, *90*, 39–40.
- [43] P. T. A. Galek, E. Pidcock, P. A. Wood, I. J. Bruno, C. R. Groom, *CrystEngComm* **2012**, *14*, 2391–2403.
- [44] a) P. T. A. Galek, L. Fábíán, W. D. S. Motherwell, F. H. Allen, N. Feeder, *Acta Crystallogr. Sect. B* **2007**, *63*, 768–782; b) P. T. A. Galek, L. Fábíán, F. H. Allen, *CrystEngComm* **2010**, *12*, 2091–2099; c) P. T. A. Galek, L. Fábíán, F. H. Allen, *Acta Crystallogr. Sect. B* **2010**, *66*, 237–252.
- [45] P. T. A. Galek, F. H. Allen, L. Fábíán, N. Feeder, *CrystEngComm* **2009**, *11*, 2634–2639.
- [46] a) A. Delori, P. T. A. Galek, E. Pidcock, W. Jones, *Chem. Eur. J.* **2012**, *18*, 6835–6846; b) A. Delori, P. T. A. Galek, E. Pidcock, M. Patni, W. Jones, *CrystEngComm* **2013**, *15*, 2916–2928; c) M. Majumder, G. Buckton, C. Rawlinson-Malone, A. C. Williams, M. J. Spillman, E. Pidcock, K. Shankland, *CrystEngComm* **2013**, *15*, 4041–4044.
- [47] O. Korb, P. A. Wood, *Chem. Commun.* **2010**, *46*, 3318–3320.
- [48] P. A. Wood, T. S. G. Olsson, J. C. Cole, S. J. Cottrell, N. Feeder, P. T. A. Galek, C. R. Groom, E. Pidcock, *CrystEngComm* **2013**, *15*, 65–72.
- [49] T. Gelbrich, D. S. Hughes, M. B. Hursthouse, T. L. Threlfall, *CrystEngComm* **2008**, *10*, 1328–1334.
- [50] <http://www.ccdc.cam.ac.uk/Community/Requestastructure/pages/Requestastructure.aspx>.
- [51] a) <http://www.crystallography.net/>; b) S. Grazulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, A. Le Bail, *J. Appl. Crystallogr.* **2009**, *42*, 726–729; c) <http://wwmm.ch.cam.ac.uk/crystaleye/index.html>; d) N. Day, J. Downing, S. Adams, N. W. England, P. Murray-Rust, *J. Appl. Crystallogr.* **2012**, *45*, 316–323.
- [52] F. H. Allen, R. Taylor, *Chem. Commun.* **2005**, 5135–5140.